Mathematical Modelling, Vol. 6, pp. 157–171, 1985
Printed in the U.S.A.

0270–0255/85 $3.00 + .00

DETERMINISTIC REGRESSION MODELS FOR PREDICTION AND CONTROL†

E. A. GALPERIN

Département des mathématiques
Université du Québec à Montréal
C.P. 8888, Succ. A, Montréal, Québec, Canada

Communicated by Xavier J. R. Avula

(*Received May* 1984)

Abstract—Given discrete observations of the input and output values over a period of past history of an unknown controlled process, a minimum order linear stationary difference equation (predictor-controller) is sought which reproduces data in the ϵ-neighborhood of the observations and represents the class of informationally equivalent regression models for the process. The problem is formulated in R^n and in the l_∞ (Chebyshev approximation) and $l_{1,\rho}$ Banach spaces. Finite linear programming methods are applied to develop effective procedures for model identification.

1. INTRODUCTION

Given a sequence of observed data, one usually assumes a linear stationary model

$$\frac{dx}{dt} = A(p)x, \qquad x(t_0) = x_0(p), \qquad t \geq t_0, \tag{1}$$

$$y(t) = h(p)x, \qquad x \in R^m, \qquad y \in R^1, \tag{2}$$

with undetermined parameters $(p_1, \ldots, p_k) = p$ to be found by fitting the vector-function $x(t, p) \in R^m$ to observed data $y_1, \ldots, y_s$; $y_i = y(t_i)$. For doing that, the least squares method is frequently used. If a model is not known a priori (as it is known, e.g. in mechanics), then its order and structure depend on the experience of the researcher, and a model is meant successful if iterations converge to a unique vector-parameter p^* robustly under variations of the initial guess p_0. In many cases this technique serves well; however, it is easy to verify that such robust convergence is not sufficient for a model to be adequate. In complex cases it is important to be able, prior to building a model, to check whether or not a model exists and, if it does, to determine the model and time intervals on which it exists.

Let us assume that observations are evenly spaced. Then (1) allows the exact sample-data representation which is stationary too:

$$x_{n+1} = Fx_n, \qquad F = \exp(A\,\Delta t) = \sum_{k=0}^{\infty} \frac{A^k\,\Delta t^k}{k!}, \tag{3}$$

$$y_n = hx_n, \qquad x_n \in R^m, \qquad y_n \in R^1, \qquad n = 0, 1, \ldots. \tag{4}$$

† This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant A3492.

Here $x_n = x(t_n)$; $t_n = t_0 + n\,\Delta t$, $\Delta t = \text{const}$; $y_n = y(t_n)$ are observations. The values of A, h, x_0, hence, F and x_n, are not known. However, $F = \text{const}$, if $A = \text{const}$ and $\Delta t = \text{const}$.

Discrete system (3), (4) yields exact values $x(t)$, $y(t)$ of the system (1), (2) at times $t_n = t_0 + n\,\Delta t$. If one takes the first two terms of the series in (3), then the system (3) with the matrix $F^* = I + \Delta t A$ (I = unit matrix) gives a discrete approximation to the system (1). The values of its state vector x_n^* do not coincide with $x(t)$ at times t_n but tend to $x(t_n)$ as $\Delta t \to 0$. So, if only discrete observations are available, it giving the only basis for validation, the discrete model (3), (4) gives complete and exact description of the behavior of the system (1) in regard to the information contained in the available discrete observations.

It is known [1, 2] that a linear stationary model (1)–(2), or (3)–(4), of the order m exists if and only if the observations $y_0, y_1, \ldots$ satisfy a linear stationary difference equation

$$y_{n+r} = a_1 y_n + a_2 y_{n+1} + \cdots + a_r y_{n+r-1}; \qquad r \leqslant m, \tag{5}$$
$$a_i = \text{const}\ (i = 1, \ldots, r); \qquad n = 0, 1, 2, \ldots.$$

Now consider two finite sequences of observed data

$$y_0, y_1, \ldots, y_N \tag{6}$$

$$u_0, u_1, \ldots, u_N, \tag{7}$$

which are in the input-output relation and suppose that the input u_i is applied after the output y_i ($i = 0, 1, \ldots, N$) has been measured, the latter being just an enumeration agreement. The two sequences describe a control system or subsystem of unknown structure. If the structure is assumed to be linear, then one can write down a difference equation:

$$y_{n+r} = a_1 y_n + a_2 y_{n+1} + \cdots + a_r y_{n+r-1} + b_1 u_n + b_2 u_{n+1} + \cdots + b_r u_{n+r-1}$$
$$(n = 0, 1, \ldots, N - r). \tag{8}$$

Here r is the order; $a = (a_1, \ldots, a_r) = \text{const}$ – dynamics of the free motion system; $b = (b_1, \ldots, b_r) = \text{const}$ – dynamics of control. If the observed data (6), (7) fit to an equation of the type (8), then parameters r, a, b can be identified and the equation so obtained can serve as a predictor (the case of u_i fixed) or a controller (u_i to be chosen) for the process represented by the data (6), (7).

Introducing the r-vectors

$$\bar{x}_n^T = [y_n, y_{n+1}, \ldots, y_{n+r-1}] \qquad (^T)\text{-transpose}, \tag{9}$$

$$v_n^T = [u_n, u_{n+1}, \ldots, u_{n+r-1}], \tag{10}$$

the Eq. (8) can be written in the form of a system:

$$\bar{x}_{n+1} = A\bar{x}_n + Bv_n, \qquad \bar{x}_n \in R^r, \qquad v_n \in R^r, \tag{11}$$

$$y_n = h\bar{x}_n, \qquad y_n \in R^1, \tag{12}$$

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline a_1 & a_2 \ldots a_r \end{array}\right], \qquad B = \left[\begin{array}{c} 0 \\ \hline b_1 b_2 \ldots b_r \end{array}\right], \tag{13}$$

$$h = [1, 0, \ldots, 0]. \tag{14}$$

The model (11)–(14) is a discrete system in canonical coordinates. One may look for a model (11)–(12) in arbitrary coordinates similar to (3)–(4) which does not have the special structure (13), (14) of the matrices. As shown in [2] for $u_i = 0$ ($i = 0, 1, \ldots, N$) and can be easily generalized for $u_i \neq 0$, every such model is equivalent to an Eq. (8) with appropriate coefficients establishing an input-output relation based on the data alone.

This shows the multiplicity of models so that Eq. (8), in fact, represents a class of informationally equivalent models with respect to the observed data (6)–(7). The advantage of an Eq. (8) is that for each order r it contains the minimum number of essential parameters.

2. CLASSES OF INFORMATIONALLY EQUIVALENT MODELS

Since the problem of model building does not have a unique solution, it is natural to consider classes of informationally equivalent models.

Definition 1. Models which generate the same output values and satisfy the same structural limitations, if any, are said to be informationally equivalent with respect to the output values (observations).

In practice, output values are known (measured) with certain precision so that unknown exact value y_n^0 belongs to a neighborhood of the observation y_n, say, $y_n - \epsilon_1 < y_n^0 < y_n + \epsilon_2 (\epsilon_1 > 0, \epsilon_2 > 0)$ or, more generally, $y_n^0 \in \Omega_n(y_n)$. In this case, informationally equivalent models are all those models which generate output values in Ω_n ($n = 0, 1, 2, \ldots$). This applies also to the case of forecasting where past observations are exact and prediction does not have to be exact but approximate to some acceptable degree μ of accuracy (see the sequel).

Definition 2. A collection of informationally equivalent models is called an equivalence class of models.

Any model of the equivalence class can be taken as its representative. From now on we shall identify an equivalence class with one of its representatives, and in the discrete linear stationary case with the simplest and most convenient difference model (8). It is clear that models within an equivalence class might be quite different mathematically but they are not physically distinguishable, so they can be thought of as various representations of the phenomenon given in observations.

Example 1. Suppose we are given the sequence of observations $y_n = 1$, $n = 0, \pm 1, \pm 2, \ldots$, which are exact. Then the functions $x \equiv 1$; $x = 1 + a \sin \pi t$, $a = \text{const}$ arbitrary; $x = |\sin(\pi/2)(2t + 1)|$ considered as models reproducing the observations are all informationally equivalent for $t \in (-\infty, +\infty)$.

Example 2. Suppose we are given the function $x = t$ with t exact and x measured with the accuracy of 0.1 mm. Then the functions $x = t$; $x = t \pm \alpha$, $|\alpha| \leq 0.1$; $x = t + 0.1 \cos e^t$; $x = t + 0.1 \sin \ln|t|$, $|x(0)| \leq 0.1$ are all informationally equivalent with respect to those measurements. If $t \in [0, 1000]$, then the function $x = (1 + \epsilon)t$, $|\epsilon| \leq 10^{-4}$ is also informationally equivalent to the above mentioned functions on [0, 1000].

3. DYNAMIC AND REGRESSION MODELS

If an exact relation (5) or (8) holds, it is a *dynamic model* representable in forms (1)–(2), (3)–(4) or (11)–(12). If (8) does not hold for any r, a, b, it means that there is no linear stationary dynamic model representing the process given in observations. In this case,

instead of (8), one can write

$$\hat{y}_{n+r} = a_1 y_n + a_2 y_{n-1} + \cdots + a_r y_{n-r-1} + b_1 u_n + b_2 u_{n-1} + \cdots + b_r u_{n+r-1}$$
$$(n = 0, 1, \ldots, N - r). \quad (15)$$

$$\eta_n(r, a, b) = y_{n+r} - \hat{y}_{n+r}; \qquad a = [a_1, \ldots, a_r], \qquad b = [b_1, \ldots, b_r]. \quad (16)$$

If $\eta_n(r, a, b)$ $(n = 0, 1, \ldots)$ for certain r, a, b are all sufficiently small, then relation (15), which we shall call the *regression model* in contrast with the dynamic model (8), gives a predictor-controller for the process given in observations (6)–(7). Clearly, regression models (15) yield broader class of predictor-controllers than dynamic models. There are many approaches to mathematical modelling of processes. Usually some structural information is assumed to be known a priori [3–7]. Here we consider modelling of processes with completely unknown structure. If the observations fit into Eqs. (15)–(16) with an acceptable degree of accuracy (see the sequel), then a process possesses the property of approximate linearity in the regression sense stated above. Otherwise, the process is essentially nonlinear. No noise is explicitly considered in the Eqs. (15)–(16), although their imprecision $\eta_n \not\equiv 0$ allows for a bounded amplitude coloured noise of unknown characteristics. This is what we call a deterministic regression that differs essentially from a dynamic difference equation since at each step it starts from a number of new updated observations. In contrast with dynamic models, deterministic regression models do not possess the semigroup property and so represent a far more powerful instrument in mathematical modelling than dynamic models.

4. LEAST-SQUARES SOLUTION IN R^n

Repeatedly writing the Eq. (8) for $n = 0, 1, \ldots, N - r$, one comes to the system:

$$Y = PX, \quad (17)$$

where

$$Y = \begin{bmatrix} y_r \\ y_{r+1} \\ \cdots \\ y_N \end{bmatrix}, \qquad P = \begin{bmatrix} y_0 \cdots & y_{r-1} u_0 \cdots & u_{r-1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ y_{N-r-1} \cdots & y_{N-1} u_{N-r-1} \cdots & u_{N-1} \end{bmatrix}, \qquad X = \begin{bmatrix} a_1 \\ \cdots \\ a_r \\ b_1 \\ \cdots \\ b_r \end{bmatrix}.$$

If for some r

$$\operatorname{rank} P = \operatorname{rank} [P, Y] = k \leq 2r \quad (18)$$

then $X = [a_1, \ldots, a_r, b_1, \ldots, b_r]^T$ is determined (uniquely, if $k = 2r$, or not, otherwise). The result is a dynamic model of order r.

If for all reasonable orders r we have

$$\operatorname{rank} P < \operatorname{rank} [P, Y] \quad (19)$$

then systems (17) are all inconsistent and one can use the Penrose pseudoinverse to obtain

the minimum-norm least-squares solution [8] for a fixed r:

$$X^0 = [a_1, \ldots, a_r, b_1, \ldots, b_r]^T = P^{\#} Y. \tag{20}$$

Resulting coefficients $X^0 = [a, b]^T$ yield a regression model. As distinct from dynamic models, a regression model does not start from one single initial point but rather at each step it starts from different successive points given in observations. So the trajectory of a regression model (15) are not the observations (6), as in the case of a dynamic model, but some other values in neighborhoods of the observations (6) and such that each value is based on r immediately preceding actual observations and not on the r very first ones $y_0, \ldots, y_{r-1}$, as in the case of a dynamic model.

The optimal solution (20) gives the residual discrepancy vector

$$Z(r) = PX^0 - Y \tag{21}$$

which is minimal in the Euclidean norm

$$\| Z \| = \sum_{i=1}^{N-r} Z_i^2 = \| PX^0 - Y \| \leq \| PX - Y \| \qquad \text{for all } X \tag{22}$$

and obviously depends on the order r fixed in advance. To obtain the best least-squares fit, one should carry out further minimization with respect to r and find

$$r_0 = \arg \min_r \| Z(r) \|, \tag{23}$$

$$Z(r_0) = PX^0 - Y|_{r=r_0}. \tag{24}$$

If all

$$| Z_i(r_0) | \leq \mu, \tag{25}$$

where μ is precision of the observations (6) or an acceptable degree of accuracy in forecasting, then a regression control model is found. If there exists one $| Z_{i_0} | > \mu$, then no acceptable least-squares regression model exists for a process given in observations (6), (7).

There is an evident upper bound for "reasonable" order of a regression model. From system (17) it can be seen that for sufficiently high orders the system is always solvable, so that linear stationary dynamic models of high orders always exist even for essentially nonlinear processes and such models can be fit to any data. Moreover, validation of such models by the past history becomes impossible and the whole construction loses ground.

Appearance of high orders in an attempt to find a good fit is an indication that the underlying process is essentially nonlinear and does not admit a linear stationary model. So the second minimization (23) should be carried up to the orders $r < N/4$; see Section 6.

5. SOLUTION IN l_∞-NORM

To meet the accuracy (25), it is sufficient to require that

$$\| Z(r_0) \| \leq \mu^2. \tag{26}$$

However, this requirement is too restrictive and may lead to rejection of all models pro-

duced by the least-squares method as well as to the wrong conclusion of inexistence of a regression model. Any weaker requirement than (26) may lead to the same result after verification by (25). So improvement can be achieved by utilizing as a performance index directly the inequality (25) instead of inequalities (22), (26) in Euclidean norm. Let us denote

$$\eta_n(r, a, b) = y_{n+r} - \sum_{i=1}^{r} a_i y_{n+r-i-1} - \sum_{i=1}^{r} b_i u_{n+r-i-1}, \tag{27}$$

$$a = [a_1, \ldots, a_r], \qquad b = [b_1, \ldots, b_r], \qquad n = 0, 1, \ldots, N - r,$$

$$\xi(r) = \inf_{a,b} \| \eta(r, a, b) \|_\infty = \inf_{a,b} \max_{0 \le n \le N-r} | \eta_n(r, a, b) |. \tag{28}$$

It is clear that inequality (25) is equivalent to

$$\xi(r_0) = \min_r \xi(r) \le \mu, \tag{29}$$

where

$$r_0 = \arg \min_r \xi(r) \tag{30}$$

Thus, two problems arise:

(a) given r, find $\xi(r)$ and $[a, b] = \arg \inf \| \eta \|_\infty$ for rth order regression model, if $\xi(r) \le \mu$;

(b) if $\xi(r) > \mu$, find r^*, if it exists, such that $\xi(r^*) \le \mu$ which gives the order of an acceptable regression model.

The first problem is the Chebyshev approximation problem which can be solved by finite programming methods [9]. To reduce it to a linear programming problem, we can treat $\xi(r)$ in (28) as a lower bound of a supplementary unknown variable a_{r+1}

$$\xi(r) = \inf_{a,b} a_{r+1}(a, b). \tag{31}$$

Definition (31) together with (28) yields

$$a_{r+1} = \max_{0 \le n \le N-r} | \eta_n(r, a, b) | \ge | \eta_n(r, a, b) |. \tag{32}$$

Relations (31), (32) with $\eta_n(r, a, b)$ defined by (27) are equivalent to the linear programming problem

$$\min \xi = a_{r+1} \tag{33}$$

under the conditions (32), that is

$$a_{r+1} - \eta_n \ge 0, \qquad n = 0, 1, \ldots, N - r, \tag{34}$$

$$a_{r+1} + \eta_n \ge 0, \qquad n = 0, 1, \ldots, N - r, \tag{35}$$

$$a_{r+1} \ge 0, \tag{36}$$

where η_n should be substituted by (27). This problem always has a solution which yields

the coefficients $a_1, \ldots, a_r, b_1, \ldots, b_r$ of a regression model and its precision a_{r-1}. The precision should satisfy $a_{r-1} \leq \mu$, otherwise one should increase the order r and repeat the solution. An effective computational procedure for solving this problem is to convert it to its dual and to apply a modification of the simpex algorithm; see [10].

Let us study the function $\xi(r)$.

THEOREM 5.1. The function $\xi(r) \geq 0$ is nonincreasing and there exists $r_* < N + 1$ such that $\xi(r_*) = 0$.

Proof. Consider first the case $b = 0$ of uncontrolled systems [11]. Let us fix $a_1 = 0$ in the setting (33)–(36) and denote the minimal value of the linear form (33) under this condition by ξ^*. It is clear that $\xi^* \geq \xi(r)$ since ξ^* is the result of minimization with an additional constraint $a_1 = 0$ compatible with the original set of constraints (34)–(36). This can be done for any $r \geq 2$ and ξ^* corresponds to $r - 1$ variable coefficients so that we have from (28) with obvious change of indexing:

$$\xi^* = \xi^*(r-1) = \xi(r)\,|_{a_1=0} = \inf_{\{a_2,\ldots,a_r\}} \max_{0\leq n\leq N-r} |\,\eta_n(r; 0, a_2, \ldots, a_r)\,|$$

$$= \inf_{\{a_1,\ldots,a_{r-1}\}} \max_{1\leq n\leq N-r+1} |\,\eta_n(r-1; a_1, \ldots, a_{r-1})\,| \geq \xi(r), \qquad r \geq 2. \tag{37}$$

For $\xi(r-1)$ we have by definition (28):

$$\xi(r-1) = \inf_{\{a_1,\ldots,a_{r-1}\}} \max_{0\leq n\leq N-r+1} |\,\eta_n(r-1, a_1, \ldots, a_{r-1})\,|$$

$$= \inf_{\{a_1,\ldots,a_{r-1}\}} \max\Big(|\,\eta_0\,|, \max_{1\leq n\leq N-r+1} |\,\eta_n\,|\Big). \tag{38}$$

Let us denote

$$\Omega_0 = \{a: |\,\eta_0\,| \geq \max_{1\leq n\leq N-r-1} |\,\eta_n\,|\} \subseteq R^{r-1}, \tag{39}$$

$$\Omega = \{a: |\,\eta_0\,| < \max_{1\leq n\leq N-r-1} |\,\eta_n\,|\} \subseteq R^{r-1}, \tag{40}$$

so that $\Omega_0 \cap \Omega = \phi$ and $\Omega_0 \cup \Omega = R^{r-1}$. By virtue of (37) we have

$$\inf_{a\in\Omega_0} |\,\eta_0\,| = |\,\eta_0(a^*)\,| \geq \max_{1\leq n\leq N-r-1} |\,\eta_n(a^*)\,| \geq \xi^*, \tag{41}$$

$$\inf_{a\in\Omega} \max_{1\leq n\leq N-r+1} |\,\eta_n\,| \geq \xi^* = \inf_{a} \max_{1\leq n\leq N-r+1} |\,\eta_n\,|. \tag{42}$$

Now from (37), (38), (41), (42) it follows

$$\xi(r-1) = \inf_a \max\Big(|\,\eta_0\,|, \max_{1\leq n\leq N-r+1} |\,\eta_n\,|\Big)$$

$$= \min\Big(\inf_{a\in\Omega_0} |\,\eta_0\,|, \inf_{a\in\Omega} \max_{1\leq n\leq N-r-1} |\,\eta_n\,|\Big) \geq \min(\xi^*, \xi^*) = \xi^* \geq \xi(r) \tag{43}$$

so the first part of the theorem is correct.

To prove the second part, let us take in (15) $b_i = 0$, $n = 0$, $r = r_* = N < N + 1$. If at least one of y_s ($s = 0, 1, \ldots, N-1$) is different from zero, one can always choose

such a_i in (15) that the following equality will hold:

$$\hat{y}_N = \sum_{i=1}^{N} a_i y_{i-1} = y_N, \tag{44}$$

where y_N is the last observation. So by (16), (28) we have that

$$\xi(r_*) = \xi(N) = \inf_a | \eta_0(N, a) | = | y_N - \hat{y}_N | = 0, \qquad r_* < N + 1. \tag{45}$$

If all $y_s = 0$, $s = 0, 1, \ldots, N-1$ and $y_N \neq 0$, then whatever r, a_i, one always has $\hat{y}_N = 0$, $\xi(r) = | y_N | > 0$, $r = 1, 2, \ldots, N$; however, such a situation is of no interest since it is, in fact, a problem with one sole observation and no prediction is possible. This completes the proof for the case $b = 0$. The case $b \neq 0$ of controlled processes differs therefrom in the point that the dimension of the parameter space is doubled. However, due to the symmetry of the Eq. (15) with respect to y_i and u_i and to the fact that these data are given as certain constants, one can apply exactly the same argument. In fact, imposing two additional constraints

$$a_1 = b_1 = 0 \tag{46}$$

on the variables of the linear programming problem (33)–(36), (27), one comes to the functional ξ^* with the obvious inequality:

$$\xi^* = \xi^*(r-1) = \xi(r)\,|_{a_1 = b_1 = 0} \geq \xi(r). \tag{47}$$

Then, exactly in the same way one proves the inequality

$$\xi(r-1) \geq \xi^*(r-1) \tag{48}$$

which together with (47) completes the proof of the first part of the theorem for the case $b \neq 0$.

The second part of the theorem for this case can be formally proved by straightforward generalization of the arguments presented above. ■

The theorem has clear physical and computational significance:

(1) with the increasing order, the precision of a regression model cannot decrease;
(2) for a sufficiently high order there always exists an exact regression model $\xi(r_*) = 0$, which represents a *dynamic* model;
(3) one can take r^* suggested by physical considerations and compute $\xi(r^*)$: if $\xi(r^*) > \mu$, then it is necessary to improve a model by taking $r > r^*$; the values $r < r^*$ are not to be considered.

It is instructive that even essentially nonlinear system given by a finite number of evenly-spaced discrete observations admits a linear dynamic model of a sufficiently high order as its informationally equivalent representative. So high r_* is an indication that a system is essentially nonlinear (the converse is obviously wrong).

On the other hand, if a nonlinear systems can be approximated by a linear one, this will be clearly established by a nonincreasing order of the corresponding linear regression model up to a certain degree of accuracy which gives the precision of allowable linear approximation; see Section 10.

6. COMPUTATION OF THE LEAST UPPER BOUND FOR THE ORDERS OF AN ADEQUATE REGRESSION MODEL

Let us compute the number r_* in the Theorem 5.1. Since r_* defines the order of a dynamic model, so the computations can be carried out in any norm without difference in precision of the resulting model.

Consider the system (17). The number $N + 1$ of observations is fixed in advance while the order r can be increased. In practice, due to imprecision of data, the $(N - r) \times 2r$-matrix P has usually full column rank, if $2r < N - r$. By the same argument the $(N - r) \times (2r + 1)$-matrix $[P, Y]$ usually has also full column rank, so that system (17) is never solvable even for linear stationary processes naturally described by a linear dynamic system. This we shall call a *regular* case, otherwise a singular case.

Since $|\eta_n(r, a, b)| = |y_{n-r} - \hat{y}_{n-r}| \leq \mu$ are allowed, so in a singular case one can always change a little the observations employed in (6), (7) and accordingly decrease μ in order to obtain regular case. In this case, taking an order r such that $2r \geq N - r$, that is

$$r \geq N/3, \tag{49}$$

one obtains a system (17) with the matrix P of a full row rank for which the system is solvable and presents a dynamic model. Clearly, $r \leq N$, so that $r \in [N/3, N]$, and to avoid or decrease the number of free coefficients, it is natural to take the least integer from this interval that yields

$$r_* = \begin{cases} N/3, & \text{if 3 divides } N \\ [N/3] + 1, & \text{otherwise.} \end{cases} \tag{50}$$

In singular cases it may happen that one finds a lesser r_* than in (50). On the other hand, the case of $y_s = 0$, $s = 0, \ldots, N - 1$, $y_N \neq 0$ does not obey the Theorem 5.1. However, this case admits a trivial "model": $y_{n+1} = y_n$ $(r = 1, a_1 = 1, b_1 = 0)$ which fails at $n = N - 1$. Such a situation is of no interest since it is, in fact, a problem with one sole observation and no prediction or control is possible. Such cases are excluded: see Section 8.

The order r_* in (50) gives the least upper bound for orders in the regular case. However, a model of this order fits any data and so does not reflect the nature of a process and has no practical significance. With this model all observations are already used for identification of coefficients and none is left for validation of a model obtained. To construct a plausible model, one should leave aside a subset of observations and use them to check a model already identified.

To accomplish this task, the orders much less than r_* in (50) should be considered in order that a model be adequate. This quality can be measured by the difference

$$\nu(r) = N - 3r \tag{51}$$

which should not be less than a certain integer M assigned, e.g. on consideration of the importance of having a confident prediction. So it is necessary to allow for a trade-off between precision and validation (verifiability) of a model which should satisfy the requirements:

$$\begin{aligned} &\xi(r) \leq \mu, \\ &\nu(r) = N - 3r \geq M > 0, \end{aligned} \tag{52}$$

The pair (μ, M) characterizes the quality of a model. In the case of a dynamic model the integer M indicates precisely the number of observations set aside for validation of a model determined by earlier observations.

Now the procedure is clear. Take an integer

$$r = \left[\frac{N - M}{3}\right] \leq \frac{N - M}{3} \tag{53}$$

and solve the linear programming problem (33)–(36). If $a_{r+1} \leq \mu$, a linear predictor-controller is found. If $a_{r+1} > \mu$, weaken, if possible, the requirements (52) and repeat the solution for increased r. If $a_{r+1} > \mu$ and (52) cannot be weakened, then no adequate regression model exists.

The notion of confidence measured by the integer M is related also to the number of observations $N + 1$, so it is expedient to introduce a relative quantitative measure for the confidence (validity) of a model. Since $M < N$, one can write

$$M = \alpha N, \qquad 0 < \alpha < 1, \tag{54}$$

where α is the percentage of observations reserved for validation purposes. With this notation the relation (53) is transformed as

$$r = \left[\frac{(1 - \alpha)N}{3}\right] \leq \frac{1 - \alpha}{3} N \tag{55}$$

and the quality of a model is characterized by the pair (μ, α). For example, if one takes $\alpha = 25\%$, then the order of an adequate regression model should be $r \leq N/4$, the bound given earlier in the Section 4.

7. SOLUTION IN $l_{1,\rho}$-NORM

In some cases it is sensible to abandon the uniformity involved in the l_∞-norm (28) since past values may have less influence on future behavior than more recent factors. This can be taken into account with a positive weight sequence $\rho_n > 0$ and naturally leads to the introduction of the $l_{1,\rho}$-norm. The performance index (28) is thus replaced by

$$\xi_1(r) = \inf_{a,b} \| \eta(r, a, b) \|_{1,\rho} = \inf_{a,b} \sum_{n=0}^{N-r} \rho_n \, | \eta_n(r, a, b) | \leq \mu_1, \tag{56}$$

where

$$\mu_1 = \mu \sum_{n=0}^{N-r} \rho_n \, . \tag{57}$$

This functional presents a convex piecewise-linear program which can be reduced to a linear programming problem by introduction of $N - r + 1$ auxiliary variables $a_{r+1}, \ldots, a_{N+1}$ as upper bounds of the deviations:

$$| \eta_n(r, a, b) | \leq a_{n+r+1} \qquad (n = 0, \ldots, N - r). \tag{58}$$

The relations (56), (58) and (27) give rise to the linear programming problem

$$\min \xi_1 = \sum_{n=0}^{N-r} \rho_n a_{n+r+1} \tag{59}$$

under the conditions (58), that is

$$a_{n+r-1} - \eta_n \geq 0, \qquad n = 0, 1, \ldots, N - r, \tag{60}$$

$$a_{n+r-1} + \eta_n \geq 0, \qquad n = 0, 1, \ldots, N - r, \tag{61}$$

$$a_{n+r+1} \geq 0, \qquad n = 0, 1, \ldots, N - r, \tag{62}$$

where η_n should be substituted by (27).

Clearly, this problem always has a solution. An effective procedure is to apply the decomposition algorithm [12] to its dual; see [13].

To obtain a result analogous to Theorem 5.1, we need to introduce a notion of consistent weight sequences.

Definition. Two finite weight sequences are said to be consistent, if the smaller one coincides with the corresponding forward part of the larger one, that is $\{\rho_1, \ldots, \rho_s\}$ is consistent with $\{\rho_1^*, \ldots, \rho_k^*\}$, $k < s$, if $\rho_{k-i}^* = \rho_{s-i}$ for $i = 0, 1, \ldots, k - 1$. It simply means that while changing the order r in (56), one does not change the relative importance of the more recent observations.

THEOREM 7.1. Whatever consistent weight sequences are employed, the function $\xi_1(r)$ in (56) does not increase with r and there exists $r_0 < N + 1$ such that $\xi_1(r_0) = 0$.

Proof. It follows the same idea: take for simplicity $b = 0$, then

$$\xi_1^* = \xi_1(r)\,|_{a_1=0} = \inf_{\{a_2,\ldots,a_r\}} \sum_{n=0}^{N-r} \rho_n \,|\, \eta_n(r; 0, a_2, \ldots, a_r) \,|$$

$$= \inf_{\{a_1,\ldots,a_{r-1}\}} \sum_{n=1}^{N-r+1} \rho_{n-1} \,|\, \eta_n(r-1; a_1, \ldots, a_{r-1}) \,| \geq \xi_1(r), \qquad r \geq 2, \tag{63}$$

$$\xi_1(r-1) = \inf_{\{a_1,\ldots,a_{r-1}\}} \sum_{n=0}^{N-r+1} \rho_n^* \,|\, \eta_n \,| = \inf_{\{a_1,\ldots,a_{r-1}\}} (\rho_0^* \,|\, \eta_0 \,| + \sum_{n=1}^{N-r+1} \rho_n^* \,|\, \eta_n \,|). \tag{64}$$

Since $\rho_n^* = \rho_{n-1}(n = 1, \ldots, N - r + 1)$, $\rho_0^* > 0$, so $\xi_1(r-1) \geq \xi_1^* \geq \xi_1(r)$. The second part of the proof remains the same as in Theorem 5.1 and generalization to the case $b \neq 0$ is straightforward. ■

The algorithm consists of the same steps as in the preceding section. The requirements for an adequate model are

$$\xi_1(r) \leq \mu_1, \qquad \nu(r) = N - 3r \geq M. \tag{65}$$

8. UNIFORMITY OF CONFIDENCE PRINCIPLE

The validity of a model is based on the following obvious axiom which we call the uniformity of confidence principle.

The axiom. If a model is good for prediction of a sufficiently large number of past observations and at present there is no new factor which may essentially affect the behavior of a system, then the model is good for prediction and control of the future. ■

It should be emphasized that any approach to modelling, in addition to experimental data, requires a bit of extra information to provide for confident conclusions. This information is either assumed a priori or known from general considerations (structural prop-

erties, etc.). The richer this information, the more accurate and confident the model. The proposition of the axiom is probably the smallest bit of additional information without which no conclusions are possible, whatever rich experimental data are available.

Assumptions like the axiom are often made tacitly in modelling. We formulate it explicitly. The formulation contains some vague terms: "good for prediction," "sufficiently large," "esentially affect." This reflects the nature of the general problem in modelling and allows a researcher to make his own specification of consistency according to a concrete real life problem [3–7], [14–16].

The axiom can be rigorously formulated in many ways. One of such formulations is presented in [17] in the form of consistent pairs "process-algorithm"; the principle itself is called "transition assumption" and a theorem that the set of consistent pairs "process-algorithm" (c.f. system-model) is not empty was proved. This specification was designed and served well for the applications considered in [17]. Here it is expedient, however, to take the axiom in the form presented and leave various formulations to be made in accordance with particular problems.

If a model fails at some time despite its being good in the past, it indicates the interference of a new factor strong enough and not sufficiently represented by the model.

9. EXPERIMENTABLE AND NONEXPERIMENTABLE CONTROL MODELS

A process under study may allow experimentation, i.e. different sequences (7) of input values can be played over to obtain different output sequences (6) for the purpose of identification.

A process (system, model) represented by (8) or (15) is said to be experimentable, if one can retrace the output y_i repeatedly for different inputs $u_i^{(1)}$, $u_i^{(2)}$, . . . , etc. Technological systems, most biological processes and certain socioeconomic systems have this property. Let all $u_i = 0$. In this case the system is in free motion, Eq. (8) coincides with (5), and the above methods can be applied to determine a_i and, thus, to obtain either dynamic or regression linear model for the free system. This task is simpler since the number of parameters is r and not $2r$ as in (8).

It is clear that, if a controlled process admits a dynamic model, i.e. an exact Eq. (8), then its free motion model (5) will be determined as a dynamic model too. On the contrary, if its free motion model with all $u_i = 0$ is of regression type, then its control model (8) cannot be of dynamic type and is necessarily a regression one which automatically implies a class of informationally equivalent models. This multiplicity was noticed and illustrated by Wu and Rekasius[18, p. 503, Example 2: "the possibility of false modelling"]; however, it was inadequately interpreted. Since the Eq. (8) allows for a play between the two dynamics, one should first establish a free motion model to fix the dynamics pertaining to the process itself and not to its controls even in the case of a regression model.

Once a free motion model of either kind has been established, one can check the existence of a linear control model (8) and, if it exists, determine all coefficients b_i ($i = 1, \ldots, r$) as follows. Fix $n = n^* \in [0, N - r]$ and make r experiments while measuring $y^{(j)}_{n^*+i}$ corresponding to $u^{(j)}_{n^*-i} = c_{ij}$ ($i, j = 0, \ldots, r$) such that the determinant $| c_{ij} | \neq 0$. Then all b_i are computed from a nonsingular linear system obtained from (8) for the inputs applied. Take another appropriate sequence of inputs $u^*_{n^*+i} \neq c_{ij}$ and make one more experiment. If the process admits a linear control model (8) on the time segment $[n^*, n^* + r] \subset [0, N]$, then, whatever $u^*_{n^*+i}$ ($i = 0, \ldots, r$) applied, the equality (8) will be satisfied with the same b_i determined in the first r experiments. Otherwise, there exists $u^*_{n^*+i}$ such that (8) will be violated and this means that, although the free process is (exactly or approximately) linear, its controls are nonlinear.

Suppose that both the free process and its controls are linear, i.e. a model (8) exists.

Let n^* take the values $0, 1, \ldots, N - r$ and repeat the above experimental study under the assumption that for each n^* a linear control model (8) exists. It might happen to be one and the same model, i.e. b_i do not depend on n^*.

In this case a linear stationary control model (8) is found.

If it happens that $b_i = b_i(n^*)$ $(i = 1, \ldots, r)$, then a model (8) contains *nonstationary* controls. In this case one has to find a predictor for $b_i(n^*)$, $n^* = 0, 1, \ldots, N - r, \ldots, N, \ldots$, in order to use the model for controlling a process for $n > N - r$, i.e. beyond the given experimental time segment $[0, N]$. This task can be accomplished by the same method: fix $i = i_0 \in [1, \ldots, r]$ and consider $b_{i_0}(n^*)$ as given observations $y_s = b_{i_0}(s)$, $s = 0, 1, \ldots, N$, for which to find a model.

Once a model (8) or (15) has been found, one can use it for control purposes under the uniformity of confidence principle.

For nonexperimentable systems it is impossible to repeat the process with different controls (past history of economy, risky or costly experiments in medicine, biology or engineering). So, the information is limited and we have to infer some conclusions from the data (6), (7) alone. Thereby the dimension of the parameter space is doubled and so is the dimension of the linear programming problem corresponding to a nonexperimentable control model. Clearly, these methods are universal and can be applied to both kinds of systems.

However, for experimentable control processes it is always expedient to reduce the dimension by making use of the additional information that can be obtained by experimentation.

10. TEST EXAMPLE OF A NONLINEAR OSCILLATOR

To test the efficiency of the method and its ability to distinguish between linear and nonlinear nature of a process given in observations, the classical system of a nonlinear pendulum can be taken ([10, p. 138]).

Consider the following second order nonlinear equation and its linear approximation

$$\mathrm{d}^2x/\mathrm{d}t^2 + \omega^2 \sin x = 0, \tag{66}$$

$$\mathrm{d}^2x/\mathrm{d}t^2 + \omega^2 x = 0, \tag{67}$$

with the initial conditions: $x(0) = x_0 > 0$, $\mathrm{d}x(0)/\mathrm{d}t = 0$. Take $\omega = 2\pi$, which corresponds to the period $T = 1$ sec of a physical pendulum with the equivalent length of 25 cm and small x_0. For small x_0 the nonlinear equation (66) is usually approximated by (67) for all practical purposes. The natural questions are:

(i) whether one can recognize the second order linear stationary system by a series of discrete observations on its trajectory, and

(ii) for how big $x_0 > 0$ the nonlinear nature of (66) reveals itself in increasing order of would-be linear regression models of (66).

For this study both Eqns. (66) and (67) were integrated with the same initial conditions by the Runge–Kutta method with the precision of Δt^5. The interval of integration was $[0, 2]$ of the length $2T = 2$ sec (two oscillations) and the increments were $\Delta t_1 = 0.025$ and $\Delta t_2 = 0.05$. Accordingly, numbers N of observations and precisions were obtained (Table 1).

Table 1

Δt	N	Precision
0.025	80	10^{-8}
0.05	40	3×10^{-7}

Table 2. Nonlinear system (66) with $\Delta t = 0.025$ and $N = 80$

x_0 rad	$k = 2$	3	4	5	6
1.7	2	4	4	6	11
1.5	2	4	4	6	8
1.3	2	3	4	6	8
1.1	2	3	4	5	6
0.9	2	2	4	4	6
0.7	2	2	4	4	6
0.5	2	2	3	4	4
0.3	2	2	2	4	4
0.1	2	2	2	2	2

Table 3. Nonlinear system (66) with $\Delta t = 0.05$ and $N = 40$

x_0 rad	$k = 2$	3	4	5	6
1.7	4	6	8	11	16
1.5	3	4	6	10	12
1.3	2	4	6	8	12
1.1	2	4	6	6	11
0.9	2	4	4	6	9
0.7	2	3	4	6	6
0.5	2	2	4	4	6
0.3	2	2	2	4	4
0.1	2	2	2	2	4

Discrete data obtained by integration of (66), (67) were taken as initial observations $y_0, y_1, \ldots, y_N$ and to these data the PASCAL program implementing Procedure 2 of the l_∞-solution (see [10]) was applied to retrieve the orders of possible regression models to reproduce the data with certain accuracy. The computational experiment was carried out for various initial conditions x_0 and different preassigned accuracies μ of forecasting in (29). Tables 2 and 3 show the order r of a model (5), (16), (28), (29) as a function of x_0 and $\mu = 10^{-k}$ for $k = 2, 3, 4, 5, 6$, in the case of free oscillations ($b = 0$ in (15), (16)).

For the linear system (67) those tables were filled with 2's only, so that for (67), $r = r(x_0, \mu) = 2 = \text{const}$ which, in fact, gives a dynamic model since $\mu = 10^{-6}$ is comparable with the precision of the Runge–Kutta computations (Table 1).

These results show the effectiveness of regression models. For the nonlinear system (66) the increasing orders correspond to nonexistence of a linear regression model. Tables 2 and 3 also show the possibility of the linear approximation (67) for the nonlinear system (66) for small oscillations whose amplitude x_0 depends on the accuracy μ and on the time increment Δt of discrete trajectories of (66), (67) taken for comparison.

11. CONCLUSIONS

Deterministic regression control models are introduced for prediction and control of processes of unknown structure given by a sequence of evenly-spaced discrete observations. Experimentable and nonexperimentable systems are considered. Classes of informationally equivalent models are studied and distinction between dynamic and deterministic regression models is emphasized. The order-precision interrelation inherent to regression models is studied to provide for the best possible quality and validation of a model.

Different methods including finite linear programming methods are proposed for identification of deterministic regression control models. The regression models obtained are suboptimal in the sense of (52) or (65). The solution is unique up to the informational equivalence defined above. The algorithms always end up with a model or with the assertion of its nonexistence under the conditions (52) or (65). These conditions and the axiom in Section 8 represent natural requirements in a situation with essentially incomplete information given in a sequence of discrete observations. Such problems have become frequent in practice and the conditions (52) or (65) present a means to handle them. The methods proposed can also be employed to distinguish between linear and nonlinear nature of a process given in observations. These methods are effective for prediction based on a time series and for control of systems with unknown dynamics.

REFERENCES

1. R. C. K. Lee, Optimal estimation, identification and control. Resarch Monograph No. 28, MIT Press, MA (1964).
2. E. A. Galperin and A. Zinger, Stationarity conditions for linear models. *Int. J. Math. Modelling*, **4**, 501–514 (1983).
3. M. Berman and R. Schoenfeld, Invariants in experimental data on linear kinetics and the formulation of models. *J. Appl. Phys.* **27**, 1361–1370 (1956).
4. R. Bellman and K. J. Astrom, On structural identifiability. *Math. Biosci.* **7**, 329–339 (1970).
5. S. I. Rubinow and A. Winzer, Compartment analysis: An inverse problem. *Math. Biosci.* **11**, 203–247 (1971).
6. E. Walter, G. Le Cardinal and P. Bertrand, On the identifiability of linear state systems. *Math. Biosci.* **31**, 131–141 (1976).
7. K. J. Astrom and P. Eykhoff, System identification—A survey. *Automatica* **7**, 123–162 (1971).
8. A. Ben Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*. John Wiley & Sons, New York, 1974.
9. S. I. Zukhovitskiy and L. I. Avdeyeva, *Linear and Convex Programming*. W. B. Saunders Co., Philadelphia/London (1966).
10. E. A. Galperin and R. Labonté, Two algorithms for solution of the discrete stationary linear predictor problem, I. *Proc. of the Third Inter. Symp. on Large Engineering Systems*, St. John's, Newfoundland, Canada (1980), pp. 133–138.
11. E. A. Galperin, Regression models for prediction and control. *Proc. of the 8th IFAC World Congress*, Vol. VI, Kyoto, Japan (1981), pp. 119–124.
12. G. B. Dantzig, *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ (1963).
13. E. A. Galperin and R. Labonté, Two algorithms for solution of the discrete stationary linear predictor problem, II. *Proc. of the 1980 Joint Automatic Control Conference*, San Francisco, CA, Vol. II, TP6G (1980).
14. L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Systems* **1**, No. 1, 3–28 (1978).
15. L. A. Zadeh, Possibility theory and soft data analysis. Memorandum No. UCB/ERL M79/66, University of California, Berkeley, CA (1979).
16. E. Sanchez, Fuzzy Logics with Application to Medical Diagnosis. *Proc. of the 1980 Joint Automatic Control Conference*, San Francisco, CA, Vol. II, FA10-D (1980).
17. E. A. Galperin, An Algorithm for Long-Term Prognosis of Deterministic Time Series. *Proc. of the First Inter. Conference on Mathematical Modelling*, Univ. of Missouri-Rolla, St. Louis, MO, Vol. I (1977), pp. 437–446.
18. Y. C. Wu and Z. V. Rekasius, Deterministic identification of linear dynamic systems. *IEEE Trans. Autom. Control AC-25*, No. 3, 501–504 (1980).